

Pengelompokan Lagu Populer untuk Musik Gym Menggunakan Metode K-Means Clustering

Pande Nyoman Weda Wesnawa^{a1}, Made Agung Raharja

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹wesnawa.2208561021@student.unud.ac.id
made.agung@unud.ac.id

Abstract

Music streaming has emerged as the primary mode for individuals to enjoy music while exercising at the gym. Spotify, among the largest music streaming platforms, surveyed 2,000 gym users in the US, revealing that 82% utilize Spotify during workouts. Studies indicate music significantly influences workout quality. This study aims to cluster popular Spotify songs of 2023 using K-Means based on audio attributes like tempo, energy, and danceability. Data sourced from Kaggle's 2023 Spotify dataset underwent preprocessing. Utilizing the Elbow method, optimal cluster count determination yielded two clusters: one apt for gym use and another unsuitable. Out of 954 songs, 72.3% were gym appropriate. Visualizations via pie charts and 3D scatter plots depicted clusters based on BPM, energy, and danceability. Purity evaluation scored 1.0, ensuring accurate cluster formation. This research aids gym proprietors in crafting strategies to select motivating music, enhancing members' workout experiences.

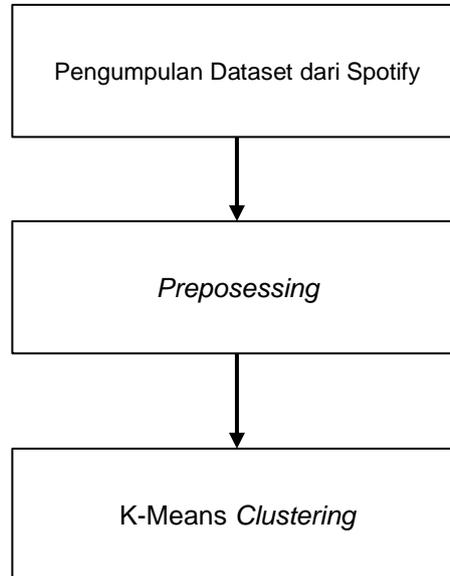
Keywords: Spotify, K-Means, Gym, Purity, Popular Song, Music Information Retrieval

1. Pendahuluan

Di era modern, *streaming* musik telah menjadi salah satu cara utama di mana individu mendengarkan musik, termasuk saat mereka melakukan aktivitas fisik seperti berolahraga gym. Dalam konteks ini, pemilihan lagu yang tepat dapat mempengaruhi motivasi dan kinerja selama latihan. Spotify, sebagai salah satu platform *streaming* musik terbesar di dunia, menyediakan akses ke berbagai jenis lagu dari berbagai genre. Pada tahun 2023, Spotify telah menjadi salah satu sumber utama musik bagi pengguna yang melakukan aktivitas gym di tempat *fitness*. Spotify melakukan survei terhadap 2.000 pengguna gym di Amerika Serikat dan menemukan bahwa 82% di antaranya menggunakan Spotify untuk mendengarkan musik saat berolahraga [1]. Pengelompokan lagu-lagu populer untuk musik gym menjadi penting untuk memahami tren dan preferensi pendengar. Metode K-Means merupakan salah satu teknik analisis klaster yang dapat digunakan untuk mengelompokkan lagu-lagu berdasarkan atribut seperti tempo, *energy*, dan *danceability*. Tempo musik yang cepat dapat meningkatkan performa olahraga dengan meningkatkan denyut jantung.[2] Energi adalah ukuran dari 0,0 hingga 1,0 dan mewakili ukuran persepsi intensitas dan aktivitas. Biasanya, trek yang energik terasa cepat, keras, dan berisik.[3] Terdapat sebuah penelitian menunjukkan bahwa musik memiliki pengaruh positif yang signifikan dalam meningkatkan performa pada latihan beban untuk meningkatkan kekuatan otot bahu dan punggung pada *member* FIK *fitness Center* di tahun 2018/2019 serta musik dengan bit 145-160 bpm terbukti menimbulkan perasaan seseorang di rentan usia muda 17-25 tahun mengalami peningkatan perasaan / *mood* menjadi semangat bergairah melakukan latihan beban [4]. Tujuan dari penelitian ini adalah untuk mengelompokkan lagu yang populer di Spotify tahun 2023 agar memudahkan dalam mengetahui lagu apa saja yang cocok untuk orang yang sering latihan olahraga di gym, serta memungkinkan untuk strategi bisnis untuk pemilik usaha *fitness* untuk membuat anggotanya menjadi nyaman latihan di tempat *fitness*.

2. Metode Penelitian

Penelitian ini meliputi beberapa langkah, yaitu pengumpulan data, *preprocessing*, dan *clustering*. Setiap tahap dalam penelitian ini memiliki fungsinya masing-masing, dan setiap tahap saling terhubung satu sama lain. Untuk memperjelas, setiap langkah akan dijelaskan sebagai berikut.



Gambar 1. Tiga Langkah Penelitian

2.1 Pengumpulan Data

Pada tahap dimulai dengan mengumpulkan lagu populer di Spotify pada tahun 2023, dataset lagu didapatkan dari kaggle. Dataset yang didapatkan berupa list lagu yang populer di tahun 2023, beserta audio fitur pada setiap lagu. Dataset yang digunakan bukan berupa audio melainkan berupa angka yang sebelumnya sudah di preprocessing terlebih dahulu oleh si pemilik dataset. Berikut adalah atribut dari datasetnya

Tabel 1. Atribut Pada Lagu

No	Nama Atribut	Keterangan
1	track_name	Nama lagu
2	artist(s)_name	Nama-nama artis yang terlibat dalam lagu.
3	artist_count	Jumlah artis yang terlibat dalam lagu.
4	released_year	Tahun rilis lagu.
5	released_month	Bulan rilis lagu.
6	released_day	Hari rilis lagu.
7	in_spotify_playlists	Lagu ini ada dalam playlist Spotify
8	in_spotify_charts	Lagu ini ada dalam chart Spotify
9	streams	Jumlah streaming lagu
10	in_apple_playlists	Lagu ini ada dalam playlist Apple Music
11	in_apple_charts	Lagu ini ada dalam chart Apple Music
12	in_deezer_playlists	Lagu ini ada dalam playlist Deezer

No	Nama Atribut	Keterangan
13	in_deezer_charts	Lagu ini ada dalam chart Deezer
14	in_shazam_charts	Lagu ini ada dalam chart Shazam
15	bpm	Beat per menit (tempo) lagu
16	key	Kunci musik utama lagu.
17	mode	Mode musik (Mayor atau Minor).
18	danceability_%	Tingkat kecocokan lagu untuk menari, dalam persentase.
19	valence_%	Tingkat positività lagu, dalam persentase.
20	energy_%	Tingkat persepsi intensitas dan aktivitas. Biasanya, trek yang energik terasa cepat, keras, dan berisik.
21	acousticness_%	Tingkat akustik lagu, dalam persentase.
22	instrumentalness_%	Tingkat instrumental lagu, dalam persentase.
23	liveness_%	Tingkat kehidupan kinerja lagu, dalam persentase.
24	speechiness_%	Tingkat kecenderungan lagu terhadap ucapan daripada musik, dalam persentase.

2.2 Preprocessing

Pada tahapan *preprocessing* ini tahapan pertama yang dilakukan adalah transformasi atau menyeleksi atribut apa saja yang akan dipakai, karena penelitian ini berfokus kepada fitur audio dari lagu maka atribut lain seperti tahun rilis lagu, nama lagu, nama artis, dan atribut lainnya yang tidak diperlukan akan dihilangkan. Setelah itu dilakukan *scaling* (normalisasi) pada semua fitur audio seperti tabel di bawah

Tabel 2. Atribut Fitur Audio

No	bpm	danceability_%	valence_%	energy_%	acousticness_%	instrumentalness_%	liveness_%	speechiness_%
0	0.087708	0.891096	1.600828	1.131729	0.151729	-0.188132	-0.745256	-0.618829
1	-1.089053	0.275624	0.407730	0.587654	-0.771972	-0.188132	-0.599314	-0.618829
2	0.551280	-1.092091	-0.827979	-0.681854	-0.387097	-0.188132	0.933082	-0.416966
3	1.692382	-0.818548	0.279898	0.466749	-0.618022	-0.188132	-0.526342	0.491420
4	0.765237	-0.134690	-1.211475	0.950371	-0.502559	7.307062	-0.526342	-0.416966

2.3 K-Means Clustering

K-Means merupakan salah satu algoritma yang bersifat *unsupervised learning*. K-Means memiliki fungsi untuk mengelompokkan data kedalam data *cluster*. Algoritma ini dapat menerima data tanpa ada label kategori. K-Means *Clustering* Algoritma juga merupakan metode *non-hierarchy*. Metode *Clustering* Algoritma adalah mengelompokkan beberapa data ke dalam kelompok yang menjelaskan data dalam satu kelompok memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda dengan data yang ada di kelompok lain. *Clustering* Algoritma (K-Means) memiliki tujuan untuk meminimalisasikan fungsi objektif yang telah di set dalam proses *clustering*. Tujuan tersebut dilakukan dengan cara meminimalkan variasi data yang ada di dalam *cluster* dan memaksimalkan variasi data yang ada di *cluster* lainnya. Tahap *clustering* K-Means diawali dengan menentukan jumlah *cluster* (k) untuk mengklasifikasikan data uji sebagai data masukan yang dikelompokkan menjadi beberapa k *cluster*. Misalnya, jika ada 900 baris data dalam kumpulan data, dan perlu membentuk 2 klaster, yakni cocok dan tidak cocok. Algoritma ini kemudian mengambil 5 baris data secara berurutan dari dataset untuk membentuk *cluster* awal. Langkah-langkah *clustering* dengan metode K-Means adalah sebagai berikut

- a. Pemilihan jumlah kluster k
- b. Ada berbagai metode yang dapat digunakan untuk menginisialisasi pusat kluster, namun salah satunya adalah dengan pendekatan acak. Pusat kluster diberi nilai awal yang ditentukan secara acak.
- c. Setiap data atau objek kemudian dialokasikan ke kluster terdekat. Kedekatan antara dua objek ditentukan berdasarkan jarak antara keduanya. Begitu pula, kedekatan data dengan kluster tertentu ditentukan oleh jarak antara data tersebut dengan pusat kluster. Pada tahap ini, perhitungan jarak antara setiap data dan setiap pusat kluster diperlukan. Jarak terjauh antara suatu data dan pusat kluster akan menentukan keanggotaan data tersebut dalam kluster tertentu. Perhitungan jarak antara data dan pusat kluster biasanya menggunakan teori jarak *Euclidean*, yang dirumuskan sebagai berikut:

$$D_{i,j} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Dimana:

$D_{i,j}$ = Jarak *Euclidean* antara titik ke-i dan titik ke-j
 $X_{1i}, X_{2i}, \dots, X_{ki}$ = Koordinat dari titik ke-i dalam ruang berdimensi k
 $X_{1j}, X_{2j}, \dots, X_{kj}$ = Koordinat dari titik ke-j dalam ruang berdimensi k

- d. Perhitungan ulang pusat kluster dilakukan dengan menggunakan anggotaan kluster yang saat ini ada. Pusat kluster adalah nilai rata-rata dari seluruh data atau objek yang termasuk dalam suatu kluster tertentu. Selain *mean*, *median* kluster juga bisa dipertimbangkan karena mean bukanlah satu-satunya metrik yang mencerminkan pusat kluster.
- e. Setelah itu, setiap objek ditetapkan kembali menggunakan pusat kluster yang baru dihitung. Jika pusat kluster tidak mengalami perubahan, proses pengelompokan selesai. Jika pusat kluster berubah, maka proses akan kembali ke langkah ketiga hingga pusat kluster stabil.

2.4 Purity

Purity adalah ukuran yang digunakan untuk menentukan seberapa murni suatu kluster dalam mengandung data dari satu kelas saja. Cara menghitungnya adalah sebagai berikut: Untuk setiap kluster, hitung jumlah titik data yang berasal dari kelas yang paling dominan dalam kluster tersebut. Kemudian, jumlahkan semua nilai tersebut dari semua kluster dan bagi dengan jumlah total titik data. Secara formal, purity dapat didefinisikan sebagai berikut jika kita memiliki himpunan kluster M dan himpunan kelas D , yang keduanya membagi N titik data [5].

$$\frac{1}{N} \sum_{m \in M} \max_{d \in D} |m \cap d| \quad (2)$$

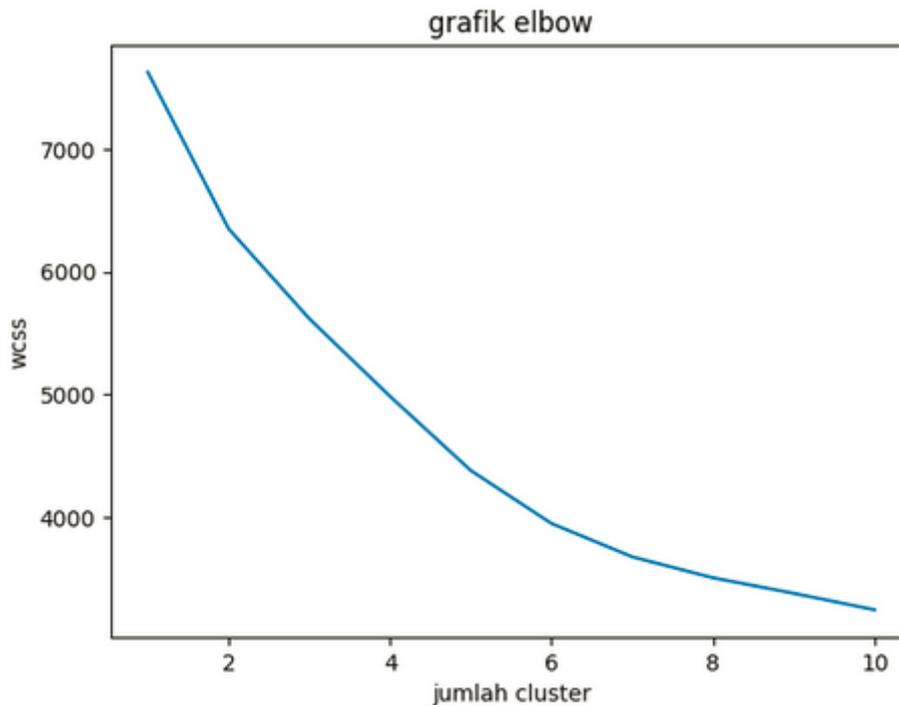
Dimana:

N = Jumlah total dalam himpunan M
 M = Himpunan elemen-elemen m
 D = Himpunan elemen-elemen d
 $\sum_{m \in M}$ = Penjumlahan untuk setiap elemen m dalam himpunan M
 $\max_{d \in D}$ = Mencari nilai maksimum dari semua d dalam himpunan D
 $|m \cap d|$ = Jumlah elemen yang sama antara m dan d .

3. Hasil dan Diskusi

Pada penelitian ini untuk menentukan jumlah *cluster* yang optimal dalam analisis K-Means *clustering* dengan menggunakan metode Elbow. Dengan menghitung WCSS untuk berbagai

jumlah *cluster* (dalam hal ini 1 hingga 10), kita dapat membuat plot WCSS versus jumlah *cluster*. Kemudian kita mencari titik elbow (sudut) dalam plot tersebut, yang menunjukkan jumlah *cluster* yang optimal. Titik elbow adalah titik di mana penurunan WCSS mulai melambat secara signifikan, menunjukkan bahwa menambahkan lebih banyak *cluster* tidak memberikan peningkatan signifikan dalam representasi data.



Gambar 2. Grafik elbow

Hasil penelitian menggunakan algoritma K-Means menghasilkan dua kluster: kluster 1 mencakup lagu-lagu yang sesuai untuk diputar selama latihan gym, sedangkan kluster 2 berisi lagu-lagu yang kurang cocok untuk situasi tersebut. Ditemukan sebanyak 694 lagu yang sesuai untuk latihan gym dan 259 lagu yang kurang sesuai dari total 954 lagu, sebagaimana terlihat dalam gambar 2, gambar 3, dan gambar 4 di bawah ini.

```
cluster
0    694
1    259
Name: count, dtype: int64
```

Gambar 2. Jumlah Pemetaan Kluster

```
[23]: c1us_0
```

track_name	artist(s)_name	artist_count	released_year	released_month	released_day	in_spotify_playlists	in_spotify_charts	streams	in_apple_playlists	in
Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	2	2023	7	14	553	147	141381703		43
LALA	Myke Towers	1	2023	3	23	1474	48	133716286		48
Cruel Summer	Taylor Swift	1	2019	8	23	7858	100	800840817		116
Sprinter	Dave, Central Cee	2	2023	6	1	2186	91	183706234		67
Ella Baila Sola	Eslabon Armado, Peso Pluma	2	2023	3	16	3090	50	725980112		34
--	--	--	--	--	--	--	--	--	--	--
Broke Boys	Drake, 21 Savage	2	2022	11	4	1060	0	106249219		3
The Great War	Taylor Swift	1	2022	10	21	1274	0	181382590		1
A Veces (feat. Feid)	Feid, Paulo Londra	2	2022	11	3	573	0	73513683		2
En La De Ella	Feid, Sech, Jhayco	3	2022	10	20	1320	0	133895612		29
Alone	Burna Boy	1	2022	11	4	782	2	96007391		27

689 rows = 24 columns

Gambar 3. Hasil Kluster 1

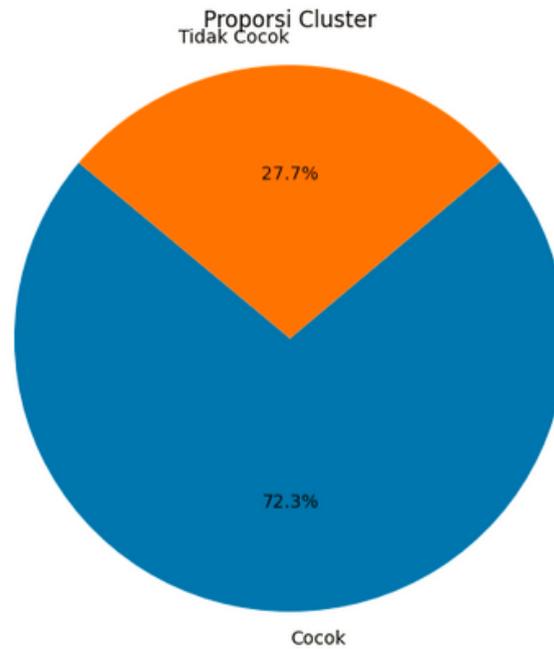
```
[24]: c1us_1
```

track_name	artist(s)_name	artist_count	released_year	released_month	released_day	in_spotify_playlists	in_spotify_charts	streams	in_apple_playlists	i
vampire	Olivia Rodrigo	1	2023	6	30	1397	113	140003974		94
WHERE SHE GOES	Bad Bunny	1	2023	5	18	3133	50	303236322		84
Daylight	David Kushner	1	2023	4	14	3528	98	387570742		80
What Was I Made For? [From The Motion Picture "Barbie"]	Billie Eilish	1	2023	7	13	873	104	30546883		80
I Wanna Be Yours	Arctic Monkeys	1	2013	1	1	12859	110	1297026226		24
--	--	--	--	--	--	--	--	--	--	--
Labyrinth	Taylor Swift	1	2022	10	21	1597	0	187339835		6
Sweet Nothing	Taylor Swift	1	2022	10	21	1747	0	186104310		9
Con La Brisa	Ludwig Goransson, Foudequish	2	2022	11	4	486	0	71095708		8
My Mind & Me	Selena Gomez	1	2022	11	3	953	0	91473363		61
Bigger Than The Whole Sky	Taylor Swift	1	2022	10	21	1180	0	121871870		4

264 rows = 24 columns

Gambar 4. Hasil Kluster 2

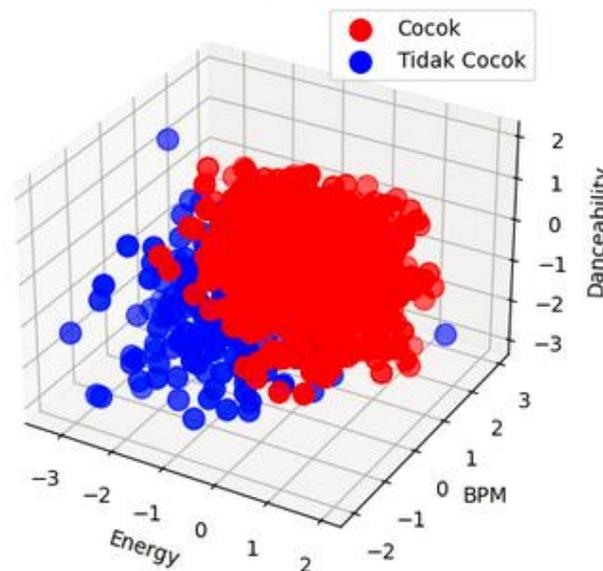
Agar memudahkan analisis, diagram *pie* menampilkan hasil kluster. Dari diagram tersebut, terlihat bahwa 72,3% lagu cocok untuk digunakan selama latihan gym, sementara 27,7% lagu tidak cocok untuk aktivitas tersebut. Berikut ini adalah gambar dari diagram *pie*.



Gambar 5. Proposi *Cluster*

Dalam mempermudah memvisualisasikan data penelitian ini menggunakan scatter plot 3D yang menampilkan dua *cluster* data berdasarkan BPM, *Energy* dan *Danceability*, dengan titik merah untuk data yang diberi label cocok dan titik biru untuk data yang diberi label tidak cocok.

Cluster Berdasarkan Energy, BPM, dan Danceability



Gambar 6. Pesebaran Data *Cluster*

Penelitian ini memanfaatkan metode purity untuk mengevaluasi jumlah dan kualitas kluster yang terbentuk. Hasil analisis menunjukkan bahwa purity score mencapai 1.0, mengindikasikan bahwa kluster yang terbentuk sudah sesuai dan akurat.

```
purity_value = purity(data_scaled['cluster'], y_kmeans)
print("Purity score:", purity_value)

Purity score: 1.0
```

Gambar 7. Nilai Purity

4. Kesimpulan

Kesimpulan penelitian ini menunjukkan data lagu yang populer di spotify pada tahun 2023 bisa dikelompokkan menjadi 2 kluster berdasarkan cocok dan tidak cocoknya suatu lagu dipergunakan untuk latihan gym. Kluster pertama yang berjumlah 694 lagu dan kluster kedua 259 lagu dari jumlah total lagu yakni 954 lagu. *Clustering* yang dilakukan telah diuji menggunakan purity dan mendapatkan nilai 1.0 yang artinya persebaran data yang didapatkan dari hasil kluster sudah bagus. Penelitian ini mengelompokkan lagu-lagu populer di Spotify tahun 2023 untuk menentukan lagu yang cocok digunakan saat berolahraga gym. Menggunakan teknik K-Means clustering, lagu-lagu diidentifikasi berdasarkan atribut audio seperti bpm/tempo, *energy*, dan *danceability*. Dataset diperoleh dari Kaggle dan telah dilakukan *preprocessing* sebelumnya. Metode Elbow menunjukkan bahwa dua *cluster* adalah optimal, satu untuk lagu yang cocok digunakan di gym (72,3% dari 954 lagu) dan satu lagi untuk yang kurang cocok (27,7%). Visualisasi hasil menggunakan diagram pie dan scatter plot 3D memudahkan analisis pada penelitian ini. *Purity score* yang mencapai 1.0 menegaskan akurasi kluster. Penelitian ini membantu pengguna dan pemilik *fitness* gym memilih musik yang tepat untuk meningkatkan motivasi dan kinerja selama latihan.

Daftar Pustaka

- [1] N. Prasetiawan, "Pengaruh Musik Pada Latihan Beban Dalam Meningkatkan Kekuatan Otot Bahu (deltoid) Dan Punggung (lastisimusdorsi) Member Fik Fitness Centertahun 2018/2019," 2020.
- [2] K. O. Dwiputra, "Manfaat Mendengarkan Musik Tempo Cepat Saat Olahraga," *KlikDokter*, Jul. 08, 2019. Accessed: May 09, 2024. [Online]. Available: <https://www.klikdokter.com/gaya-hidup/sehat-bugar/manfaat-mendengarkan-musik-tempo-cepat-saat-olahraga>
- [3] "Web API Reference," *Spotify for Developers*. <https://developer.spotify.com/documentation/web-api/reference/get-several-audio-features> (accessed May 09, 2024).
- [4] N. Prasetiawan, "Pengaruh Musik Pada Latihan Beban Dalam Meningkatkan Kekuatan Otot Bahu (deltoid) Dan Punggung (lastisimusdorsi) Member Fik Fitness Centertahun 2018/2019," 2020.
- [5] vincydesy, "Evaluation of Supervised Clustering (Purity) from scratch," *Medium*, Nov. 17, 2022. Accessed: May 24, 2024. [Online]. Available: https://medium-com.translate.google/@vincydesy96/evaluation-of-supervised-clustering-purity-from-scratch-3ce42e1491b1?_x_tr_sl=auto&_x_tr_tl=en&_x_tr_hl=en&_x_tr_hist=true